Accurate basis set truncation for wavefunction embedding

Taylor A. Barnes, Jason D. Goodpaster, Frederick R. Manby, and Thomas F. Miller

Citation: J. Chem. Phys. **139**, 024103 (2013); doi: 10.1063/1.4811112
View online: http://dx.doi.org/10.1063/1.4811112
View Table of Contents: http://jcp.aip.org/resource/1/JCPSA6/v139/i2
Published by the AIP Publishing LLC.

Accurate basis set truncation for wavefunction embedding

Taylor A. Barnes,[1] Jason D. Goodpaster,[1] Frederick R. Manby,[2] and Thomas F. Miller III[1,a)]
[1]*Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125, USA*
[2]*Centre for Computational Chemistry, School of Chemistry, University of Bristol, Bristol BS8 1TS, United Kingdom*

(Received 14 April 2013; accepted 31 May 2013; published online 8 July 2013)

Density functional theory (DFT) provides a formally exact framework for performing embedded subsystem electronic structure calculations, including DFT-in-DFT and wavefunction theory-in-DFT descriptions. In the interest of efficiency, it is desirable to truncate the atomic orbital basis set in which the subsystem calculation is performed, thus avoiding high-order scaling with respect to the size of the MO virtual space. In this study, we extend a recently introduced projection-based embedding method [F. R. Manby, M. Stella, J. D. Goodpaster, and T. F. Miller III, J. Chem. Theory Comput. **8**, 2564 (2012)] to allow for the systematic and accurate truncation of the embedded subsystem basis set. The approach is applied to both covalently and non-covalently bound test cases, including water clusters and polypeptide chains, and it is demonstrated that errors associated with basis set truncation are controllable to well within chemical accuracy. Furthermore, we show that this approach allows for switching between accurate projection-based embedding and DFT embedding with approximate kinetic energy (KE) functionals; in this sense, the approach provides a means of systematically improving upon the use of approximate KE functionals in DFT embedding. © *2013 AIP Publishing LLC*. [http://dx.doi.org/10.1063/1.4811112]

I. INTRODUCTION

The computational cost of electronic structure calculations has motivated the development of methods to partition the description of large systems into smaller subsystem calculations. Among these are the QM/MM,[1–6] ONIOM,[7,8] fragment molecular orbital (FMO),[9–15] and wavefunction theory (WFT)-in-density functional theory (DFT) embedding[16–30] approaches, which allow for the treatment of systems that would not be practical using conventional WFT approaches. In particular, WFT-in-DFT embedding utilizes the theoretical framework of DFT embedding to enable the WFT description of a given subsystem in the effective potential that is created by the remaining electronic density of the system.[16–30] We recently introduced a simple, projection-based method for performing accurate WFT-in-DFT embedding calculations[30] that avoids the need for a numerically challenging optimized effective potential (OEP) calculation[24,25,31–34] via the introduction of a level-shift operator. It was shown that this method enables the accurate calculation of WFT-in-DFT subsystem correlation energies, as well as many-body expansions (MBEs) of the total WFT correlation energy.[30]

In our original implementation, projection-based embedding was performed in the supermolecular basis, such that the embedded subsystem electronic structure calculation is performed in the atomic orbital (AO) basis set of the full system.[30] From a computational efficiency standpoint, this is not ideal. Although the embedded subsystem calculation has fewer occupied MOs than that performed over the full system, the number of virtual MOs is not reduced. The cost of traditional WFT methods typically depends more strongly on the number of virtual MOs than on the number of occupied MOs; for example, CCSD(T) method scales as o^3v^4, where o and v indicate the number of occupied and virtual MOs, respectively.[35] Truncation of the AO basis set in which the embedded subsystem is represented would lead to a reduction in the number of virtual MOs, thus significantly reducing the computational cost of the embedded subsystem calculation.

In the current work, we present a method for accurately truncating the AO basis set for embedded subsystem calculations, and we demonstrate its accuracy for both covalently and non-covalently bound systems. It is shown that this approach provides a means of controlling truncation errors and of systematically switching between existing approximate embedding methods and rigorous projection-based embedding. Furthermore, we present both embedded WFT calculations and embedded MBE (EMBE) calculations for molecular clusters and polypeptides.

II. PROJECTION-BASED EMBEDDING

We now review the projection-based embedding method,[30] which provides a rigorous framework for embedding either a WFT subsystem description in a self-consistent field (SCF) environment (WFT-in-SCF embedding) or an SCF subsystem description in an SCF environment (SCF-in-SCF embedding). The method builds upon earlier ideas to maintain orthogonality between subsystem orbitals, including frozen-core approximations,[36] the Philips-Kleinman pseudopotential approach,[37] the incremental scheme of Stoll *et al.*,[38] the region method of Mata *et al.*,[39] and Henderson's embedding scheme.[40]

a)Author to whom correspondence should be addressed: tfm@caltech.edu.

In projection-based embedding, an SCF calculation (either HF or Kohn-Sham (KS)-DFT) is first performed over the full system. The resulting set of occupied MOs, $\{\phi_i\}$, is then optionally rotated before it is partitioned into the sets $\{\phi_i\}_\mathrm{A}$ and $\{\phi_i\}_\mathrm{B}$, which correspond to subsystems A and B, respectively. These two sets of orbitals are used to construct the respective subsystem density matrices in the AO basis set, γ^A and γ^B.

In the embedded subsystem calculation, orthogonality between the subsystem MOs is enforced via the addition of a projection operator, $\mathbf{P}^\mathrm{B}$, to the subsystem A embedded Fock matrix, such that

$$\mathbf{f}^\mathrm{A} = \mathbf{h}^{\mathrm{A\ in\ B}}[\gamma^\mathrm{A}, \gamma^\mathrm{B}] + \mathbf{g}\left[\gamma^\mathrm{A}_\mathrm{emb}\right], \tag{1}$$

where the embedded core Hamiltonian is

$$\mathbf{h}^{\mathrm{A\ in\ B}}[\gamma^\mathrm{A}, \gamma^\mathrm{B}] = \mathbf{h} + \mathbf{g}[\gamma^\mathrm{A} + \gamma^\mathrm{B}] - \mathbf{g}[\gamma^\mathrm{A}] + \mu\mathbf{P}^\mathrm{B}, \tag{2}$$

$\mathbf{h}$ is the standard one-electron core Hamiltonian, $\mathbf{g}$ includes all two-electron terms, and μ is a level-shift parameter; $\gamma^\mathrm{A}_\mathrm{emb}$ is the density matrix associated with the MO eigenfunctions of $\mathbf{f}^\mathrm{A}$. The projection operator is given by

$$P^\mathrm{B}_{\alpha\beta} \equiv \langle b_\alpha| \left\{ \sum_{i\in\mathrm{B}} |\phi_i\rangle\langle\phi_i| \right\} |b_\beta\rangle, \tag{3}$$

where the b_α are the AO basis functions and the summation spans the MOs in $\{\phi_i\}_\mathrm{B}$. In the limit of $\mu \to \infty$, the MOs of subsystem A are constrained to be mutually orthogonal with those of subsystem B.[36–43] Enforcement of this orthogonality condition eliminates the need for an OEP calculation, since non-additive contributions to the kinetic energy vanish in this limit. The embedded SCF calculation using the Fock matrix in Eq. (1) is iterated to self-consistency with respect to $\gamma^\mathrm{A}_\mathrm{emb}$. The energy of the resulting SCF-in-SCF embedding calculation is then

$$\begin{aligned} E_\mathrm{SCF}\left[\gamma^\mathrm{A}_\mathrm{emb}; \gamma^\mathrm{A}, \gamma^\mathrm{B}\right] &= E_\mathrm{SCF}\left[\gamma^\mathrm{A}_\mathrm{emb}\right] + E_\mathrm{SCF}[\gamma^\mathrm{B}] + E^\mathrm{nad}_\mathrm{SCF}[\gamma^\mathrm{A}, \gamma^\mathrm{B}] \\ &\quad + \mathrm{tr}\left[\left(\gamma^\mathrm{A}_\mathrm{emb} - \gamma^\mathrm{A}\right)\left(\mathbf{h}^{\mathrm{A\ in\ B}}[\gamma^\mathrm{A}, \gamma^\mathrm{B}] - \mathbf{h}\right)\right], \end{aligned} \tag{4}$$

where E_SCF is the SCF energy and $E^\mathrm{nad}_\mathrm{SCF}[\gamma^\mathrm{A}, \gamma^\mathrm{B}]$ is the non-additive interaction energy between the densities γ^A and γ^B. The last term in Eq. (4) is a first-order correction to the difference between $E^\mathrm{nad}_\mathrm{SCF}[\gamma^\mathrm{A}, \gamma^\mathrm{B}]$ and $E^\mathrm{nad}_\mathrm{SCF}[\gamma^\mathrm{A}_\mathrm{emb}, \gamma^\mathrm{B}]$.[25] For $\mu \to \infty$, the SCF-in-SCF embedding energy is identical to the energy of the corresponding SCF calculation performed over the full system; as a result, the projection-based approach is numerically exact for SCF-in-SCF embedding calculations. In our previous work,[30] we introduced an additional perturbative correction to the SCF-in-SCF energy to account for the finite value of μ in a given computation; this correction is typically far smaller than the energy differences discussed in the current paper and is thus neglected throughout.

For the special case of DFT-in-DFT embedding, the two-electron potential terms include contributions from the electron-electron electrostatic repulsion and exchange-correlation (XC), such that

$$\mathbf{g}[\gamma^\mathrm{A} + \gamma^\mathrm{B}] = \mathbf{J}[\gamma^\mathrm{A} + \gamma^\mathrm{B}] + \mathbf{v}_\mathrm{xc}[\gamma^\mathrm{A} + \gamma^\mathrm{B}]. \tag{5}$$

The associated non-additive interaction energy is

$$E^\mathrm{nad}_\mathrm{SCF}[\gamma^\mathrm{A}, \gamma^\mathrm{B}] = J^\mathrm{nad}[\gamma^\mathrm{A}, \gamma^\mathrm{B}] + E^\mathrm{nad}_\mathrm{xc}[\gamma^\mathrm{A}, \gamma^\mathrm{B}], \tag{6}$$

where

$$J^\mathrm{nad}[\gamma^\mathrm{A}, \gamma^\mathrm{B}] = \int \mathrm{d}r_1 \int \mathrm{d}r_2 \frac{\gamma^\mathrm{A}(1)\gamma^\mathrm{B}(2)}{r_{12}} \tag{7}$$

and

$$E^\mathrm{nad}_\mathrm{xc}[\gamma^\mathrm{A}, \gamma^\mathrm{B}] = E_\mathrm{xc}[\gamma^\mathrm{A} + \gamma^\mathrm{B}] - E_\mathrm{xc}[\gamma^\mathrm{A}] - E_\mathrm{xc}[\gamma^\mathrm{B}]. \tag{8}$$

Evaluation of $J^\mathrm{nad}[\gamma^\mathrm{A}, \gamma^\mathrm{B}]$ is straightforward, and although the exact form of $E^\mathrm{nad}_\mathrm{xc}[\gamma^\mathrm{A}, \gamma^\mathrm{B}]$ is not known, approximate XC functionals are well established. Equation (6) does not include any contributions from the non-additive kinetic energy (NAKE), $T^\mathrm{nad}_\mathrm{s}[\gamma^\mathrm{A}, \gamma^\mathrm{B}]$, as this term vanishes due to the explicit mutual orthogonalization of the subsystem MOs. The special case of HF-in-HF embedding is similarly obtained by replacing the exchange-correlation potential and energy functionals, $\mathbf{v}_\mathrm{xc}[\gamma^\mathrm{A} + \gamma^\mathrm{B}]$ in Eq. (5) and $E^\mathrm{nad}_\mathrm{xc}[\gamma^\mathrm{A}, \gamma^\mathrm{B}]$ in Eq. (6), with the corresponding HF exchange terms.[30]

Projection-based embedding also allows for WFT-in-SCF embedding, in which subsystem A is treated at the WFT level and subsystem B is described at the SCF level.[30] This simply involves replacing the standard one-electron core Hamiltonian in a WFT calculation with the embedded core Hamiltonian of Eq. (2). The electronic energy from the WFT-in-SCF approach is

$$\begin{aligned} E_\mathrm{WFT}[\Psi^\mathrm{A}; \gamma^\mathrm{A}, \gamma^\mathrm{B}] &= \langle\Psi^\mathrm{A}|\hat{H}^{\mathrm{A\ in\ B}}[\gamma^\mathrm{A}, \gamma^\mathrm{B}]|\Psi^\mathrm{A}\rangle \\ &\quad + E_\mathrm{SCF}[\gamma^\mathrm{B}] + E^\mathrm{nad}_\mathrm{SCF}[\gamma^\mathrm{A}, \gamma^\mathrm{B}] \\ &\quad - \mathrm{tr}[\gamma^\mathrm{A}(\mathbf{h}^{\mathrm{A\ in\ B}}[\gamma^\mathrm{A}, \gamma^\mathrm{B}] - \mathbf{h})], \end{aligned} \tag{9}$$

where $|\Psi^\mathrm{A}\rangle$ is the embedded wavefunction from the WFT-in-SCF embedding calculation and $\hat{H}^{\mathrm{A\ in\ B}}[\gamma^\mathrm{A}, \gamma^\mathrm{B}]$ is the Hamiltonian resulting from replacing the standard core Hamiltonian with the embedded core Hamiltonian. Because the term $\mathrm{tr}[\gamma^\mathrm{A}_\mathrm{emb}(\mathbf{h}^{\mathrm{A\ in\ B}}[\gamma^\mathrm{A}, \gamma^\mathrm{B}] - \mathbf{h})]$ is included in the first term of Eq. (9), it does not appear in the last term, unlike Eq. (4).

III. AO BASIS SET TRUNCATION

A. The challenges of AO basis set truncation

Practical implementation of WFT-in-DFT embedding for large systems requires truncation of the AO basis set for the subsystem that is described at the WFT level of theory. We now illustrate the challenges of this task by analyzing the errors that arise from truncation of the AO basis set; in particular, we show that significant numerical errors can arise due to the difficulty of constructing MOs in the truncated AO basis set that are sufficiently orthogonal to the projected MOs in subsystem B.

Calculations utilizing the truncated AO basis set are referred to as truncated embedding calculations, as opposed to supermolecular embedding calculations for which the AO basis set is not truncated. Specifically, the truncated embedding calculation for subsystem A is performed within an AO basis set, $\{b_\alpha\}_\mathrm{A}$, that is a subset of the AO basis set for the full

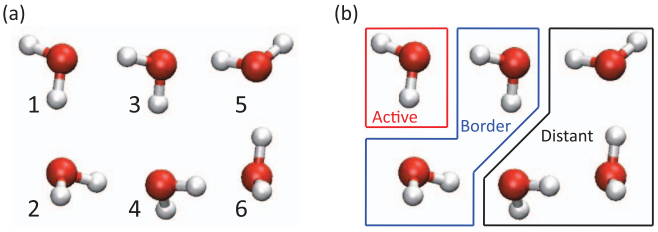

FIG. 1. (a) The BK-1 water hexamer, with molecule numbering indicated. (b) Illustration of the atom sets defined in Sec. III B, with one possible choice of the active, border, and distant atoms indicated.

system, $\{b_\alpha\}$. All calculations are performed using the implementation of projection-based embedding in the MOLPRO software package.[44]

As a starting point, we present a set of HF-in-HF supermolecular embedding calculations against which truncated embedding calculations can be compared. A closed-shell HF calculation is performed on a water hexamer in the BK-1 geometry[45] using the cc-pVDZ basis set;[46,47] all geometries employed in this study are provided in the supplementary material.[48] We number the molecules of the water hexamer as shown in Fig. 1(a). Following Pipek-Mezey localization of the canonical HF MOs,[49] subsystem partitioning is performed by assigning the five MOs with the largest Mulliken population on water molecule 1 to $\{\phi_i\}_\mathrm{A}$; the remaining MOs are assigned to $\{\phi_i\}_\mathrm{B}$. A HF-in-HF embedding calculation is then performed over a range of values for the level-shift parameter μ.

The solid line in Fig. 2(a) presents the μ-dependence of the HF-in-HF embedding error,

$$E^{\mathrm{HF}}_{\mathrm{err}} \equiv E^{\mathrm{HF}}_{\mathrm{emb}} - E^{\mathrm{HF}}_{\mathrm{full}}, \tag{10}$$

where $E^{\mathrm{HF}}_{\mathrm{emb}}$ is the energy of the HF-in-HF embedding calculation, and $E^{\mathrm{HF}}_{\mathrm{full}}$ is the energy of the HF calculation performed over the full system. As previously observed,[30] the error in the SCF-in-SCF supermolecular embedding calculations is submicrohartree and varies little with respect to μ over several orders of magnitude.

The dashed line in Fig. 2(a) shows the results of a naive HF-in-HF truncated embedding calculation, in which $\{b_\alpha\}_\mathrm{A}$ is defined to include only the AO basis functions centered on the atoms in water molecules 1, 2, and 3. Calculation of the HF MOs, $\{\phi_i\}$, and the subsystem density matrices, γ^A and γ^B, is performed in the supermolecular basis, $\{b_\alpha\}$. The embedded core Hamiltonian in Eq. (2) is initially constructed in the supermolecular basis, after which all matrix elements in $\mathbf{h}^{\mathrm{A\ in\ B}}$ that do not correspond to the truncated AO basis are discarded. The embedded calculation for subsystem A is then performed in the truncated AO basis. Unlike the supermolecular case, Fig. 2(a) illustrates that these naive truncated embedding calculations (solid) produce energies which strongly vary with respect to μ.

The dashed-dotted line and the crosses in Fig. 2(a) show the dependence of errors in the truncated embedding calculations with respect to the choice of which MOs in subsystem B are projected. In these results, the projection operator is partitioned into two parts, $P^{\mathrm{B}'}_{\alpha\beta}$ and $P^{\mathrm{B}''}_{\alpha\beta}$, each with a different level-shift parameter. The partitioned projection operators are

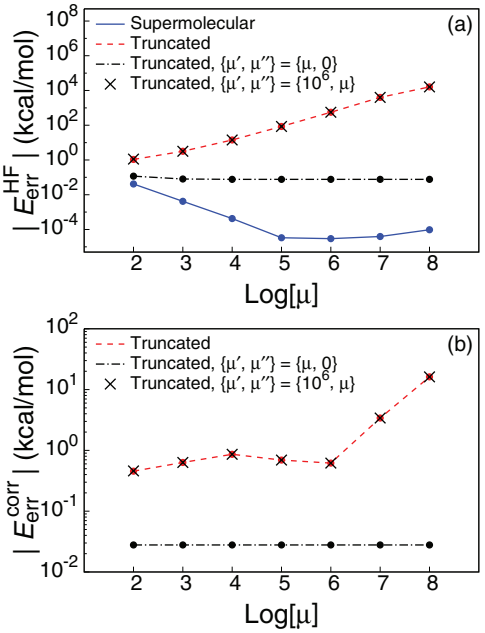

FIG. 2. (a) HF-in-HF embedding error for molecule 1 of the BK-1 water hexamer. The solid curve provides the supermolecular embedding results, while the results of naive truncation of the AO basis set are shown in the dashed curve. Also shown is the effect of partitioning the projection operator for HF-in-HF embedding in the truncated basis set, with either $\{\mu', \mu''\} = \{\mu, 0\}$ (dashed-dotted) or $\{\mu', \mu''\} = \{10^6, \mu\}$ (crosses). (b) The corresponding truncation error for the CCSD(T)-in-HF truncated embedding calculations.

defined as

$$P^{\mathrm{B}'}_{\alpha\beta} \equiv \langle b_\alpha | \left\{ \sum_{i \in \mathrm{B}'} |\phi_i\rangle\langle\phi_i| \right\} |b_\beta\rangle \tag{11}$$

and

$$P^{\mathrm{B}''}_{\alpha\beta} \equiv P^{\mathrm{B}}_{\alpha\beta} - P^{\mathrm{B}'}_{\alpha\beta}. \tag{12}$$

The summation in Eq. (11) is over the set of MOs $\{\phi_i\}_{\mathrm{B}'}$, which is a subset of $\{\phi_i\}_\mathrm{B}$. Eq. (12) corresponds to the projection of the set of MOs, $\{\phi_i\}_{\mathrm{B}''}$, that consists of all subsystem B MOs that are not included in $\{\phi_i\}_{\mathrm{B}'}$. The resulting embedded core Hamiltonian (from Eq. (2)) is

$$\mathbf{h}^{\mathrm{A\ in\ B}} = \mathbf{h} + \mathbf{g}[\gamma^\mathrm{A} + \gamma^\mathrm{B}] - \mathbf{g}[\gamma^\mathrm{A}] + \mu'\mathbf{P}^{\mathrm{B}'} + \mu''\mathbf{P}^{\mathrm{B}''}. \tag{13}$$

In these calculations, a particular MO in $\{\phi_i\}_\mathrm{B}$ is assigned to $\{\phi_i\}_{\mathrm{B}'}$ only if its combined Mulliken population on the basis functions centered on water molecules 2 and 3 is greater than 0.5, such that only the 10 (doubly-occupied) MOs in subsystem B that are localized on water molecules 2 and 3 are included. Setting μ' to a positive value while $\mu'' = 0$ corresponds to projecting only the MOs that are localized within the truncated AO basis set, $\{b_\alpha\}_\mathrm{A}$.

As illustrated by the dashed-dotted curve in Fig. 2(a), the error in the truncated embedding calculation exhibits very little dependence on μ', which suggests that the μ-dependence observed in the dashed curve is caused primarily by projection of the subsystem B MOs that are not localized within the AO basis set accessible to subsystem A. This conclusion is

also supported by the set of crosses, which shows the effect of changing μ'' while leaving μ' fixed at 10^6.

The results from Fig. 2(a) may seem counterintuitive, since the overlap between $\{\phi_i\}_{B''}$ and the truncated AO basis set is much smaller than the overlap between $\{\phi_i\}_{B'}$ and the truncated AO basis set; it might be expected that projection of the MOs in $\{\phi_i\}_{B''}$ would have little impact on the truncated embedding calculation. However, the observed behavior can be understood in terms of the difficulty of constructing MOs that are orthogonal to $\{\phi_i\}_{B''}$ within the truncated Hilbert space of subsystem A. Because the orbitals that are projected by μ'' do not strongly overlap with the basis functions accessible to subsystem A, achieving orthogonality between the subsystem A MOs and $\{\phi_i\}_{B''}$ places severe demands on the diffuse functions of the truncated AO basis set; in the supermolecular basis set, this difficulty is eliminated. For cases in which the truncated basis set is insufficiently flexible to construct MOs that are effectively orthogonal to $\{\phi_i\}_{B''}$, the error in the truncated embedding calculation increases linearly with the level-shift parameter μ''.

Figure 2(b) shows that the same trends hold for WFT-in-HF embedding. The figure plots the truncation error in the correlation energy of the WFT-in-HF embedding calculations,

$$E_{\text{err}}^{\text{corr}} \equiv E_{\text{trunc}}^{\text{corr}} - E_{\text{super}}^{\text{corr}}, \tag{14}$$

where $E_{\text{trunc}}^{\text{corr}}$ is the correlation energy of a WFT-in-HF truncated embedding calculation (i.e., the difference between the WFT-in-HF and HF-in-HF embedding energies) and $E_{\text{super}}^{\text{corr}}$ is the correlation energy of a WFT-in-HF supermolecular embedding calculation obtained with the same choices of $\{\phi_i\}_{B}$ and μ'. In the supermolecular embedding calculation, all members of $\{\phi_i\}_{B}$ are assigned to $\{\phi_i\}_{B'}$. The correlation energy is defined in the standard way,

$$E^{\text{corr}} \equiv E^{\text{WFT}} - E^{\text{HF}}. \tag{15}$$

The WFT calculations in Fig. 2(b) are performed at the CCSD(T) level of theory,[50] and the subsystems are partitioned as in the corresponding HF-in-HF embedding calculations. As observed for the HF-in-HF truncated embedding calculations, the errors of the CCSD(T)-in-HF truncated embedding calculations exhibit very little dependence on μ' and strong dependence on μ''.

Taken together, the results in Fig. 2 illustrate that significant numerical artifacts arise from the enforcement of orthogonality between the MOs of subsystem A in the truncated basis set and the MOs of subsystem B that are localized outside of the truncated AO basis set. Projection of $\{\phi_i\}_{B''}$ leads to significant errors, as well as dependence upon the level-shift parameter (Figs. 2(a) and 2(b), crosses). This problem is avoided by setting $\mu'' = 0$ in Eq. (13), resulting in truncated embedding calculations that exhibit both good accuracy and very little dependence on the remaining level-shift parameter, μ' (Figs. 2(a) and 2(b), dashed-dotted curve).

B. An improved AO basis set truncation algorithm

Incorporating the observations from Sec. III A, we now present an algorithm for AO basis set truncation in projection-based embedding that avoids dependence on the level-shift parameters and that yields controllable error with respect to the size of the truncated basis set. Truncated embedding calculations require specification of (i) the subsystem B MOs, $\{\phi_i\}_{B}$, (ii) the set of AO basis functions in which subsystem A is solved, $\{b_\alpha\}_{A}$, and (iii) the set of subsystem B MOs that are to be projected, $\{\phi_i\}_{B'}$. In the new algorithm, these specifications are made via the respective selection of (i) a set of "active atoms" that are associated with subsystem A, (ii) a set of "border atoms" that lie at the interface of subsystems A and B, and (iii) an MO overlap threshold parameter, τ.

The set of active atoms is used to determine $\{\phi_i\}_{B}$. An SCF calculation is performed over the full system using either HF theory or KS-DFT, followed by localization of the MOs; we employ the Pipek-Mezey localization method throughout this paper. An MO is assigned to $\{\phi_i\}_{B}$ if and only if the atom on which the MO has the largest Mulliken population is not an active atom. For the BK-1 water hexamer, one example of a choice of active atoms is provided in Fig. 1(b).

The set of border atoms is used to determine $\{b_\alpha\}_{A}$. Only AO basis functions centered on either an active atom or a border atom are included in $\{b_\alpha\}_{A}$. Any atom that is not assigned to either the set of active atoms or the set of border atoms is assigned to the set of "distant atoms." The special case in which no atoms are included in the set of border atoms is equivalent to using the monomolecular basis, while the special case in which no atoms are included in the set of distant atoms corresponds to using the supermolecular basis. An example of one possible choice of border atoms is given in Fig. 1(b).

The overlap threshold parameter τ is used to determine $\{\phi_i\}_{B'}$. A given MO in $\{\phi_i\}_{B}$ is assigned to $\{\phi_i\}_{B'}$ if it exhibits a combined electronic population on the border atoms, N_i, such that $|N_i| > \tau$; for the purpose of determining the electronic population on individual atoms, we employ Mulliken population analysis throughout this paper. For the special case of $\tau = 0$, all MOs in $\{\phi_i\}_{B}$ are assigned to $\{\phi_i\}_{B'}$, whereas sufficiently large values of τ correspond to assigning no MOs to $\{\phi_i\}_{B'}$.

Figs. 3(a) and 3(b) illustrate the effect of τ on the number of projected MOs and on the accuracy of HF-in-HF truncated embedding calculations, respectively. The calculations are performed using the BK-1 water hexamer geometry, and the sets of active and border atoms correspond to the case shown in Fig. 1(b). The level-shift parameters are set to $\{\mu', \mu''\} = \{10^6, 0\}$, and HF-in-HF truncated embedding calculations using the cc-pVDZ basis set (i.e., HF-in-HF/cc-pVDZ truncated embedding calculations) are performed over a range of τ. These calculations correspond to changing the number of projected MOs, while leaving the size of the truncated AO basis set unchanged. As τ approaches zero, the number of MOs in $\{\phi_i\}_{B'}$ approaches the total number of MOs in $\{\phi_i\}_{B}$ (Fig. 3(a)). As more MOs are added to $\{\phi_i\}_{B'}$, the error increases substantially (Fig. 3(b)); this is consistent with the previous observation that projection of the subsystem B MOs not localized within $\{b_\alpha\}_{A}$ results in large errors (Fig. 2, crosses). For very large values of τ, the error in Fig. 3(b) increases substantially due to "charge leakage," which is discussed later in this section and in Sec. III C.

Fig. 3(c) illustrates the sensitivity of HF-in-HF truncated embedding calculations to the size of the truncated AO

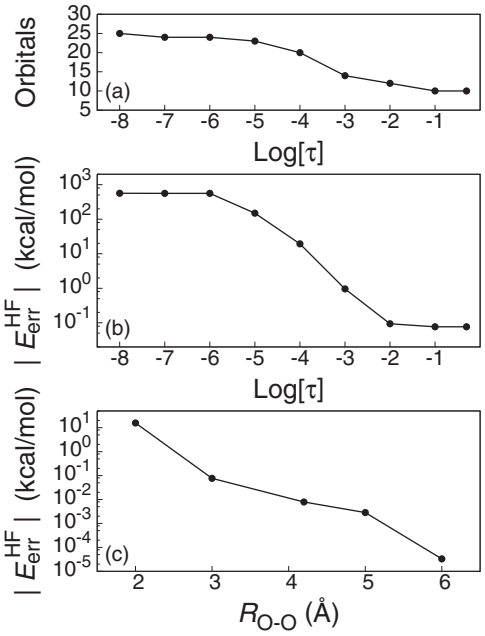

FIG. 3. (a) The number of MOs assigned to $\{\phi_i\}_{B'}$ as a function of τ for the BK-1 conformation of the water hexamer. The sets of active and border atoms correspond to the case shown in Fig. 1(b). (b) The absolute error in the HF-in-HF embedding calculation as a function of τ. The data point on the far right is equivalent to the dashed-dotted curve in Fig. 2(a) at $\mu' = 10^6$, while the data point on the far left is equivalent to the cross at $\mu'' = 10^6$. Thus changing τ corresponds to switching between the dashed-dotted curve and the set of crosses in Fig. 2(a). (c) The absolute error in the HF-in-HF embedding calculation as a function of the border atom cutoff, R_{O-O}.

basis set. The calculations use the set of active atoms indicated in Fig. 1(b) and $\{\mu', \mu'', \tau\} = \{10^6, 0, 0.5\}$. The set of border atoms for each calculation is determined through the use of a cutoff parameter, R_{O-O}. If the oxygen atom of a particular water molecule is within a distance R_{O-O} of an active oxygen atom, all atoms in that water molecule are included in the set of border atoms; the set of border atoms for each value of R_{O-O} in Fig. 3(c) is indicated in Table I. Fig. 3(c) illustrates that the truncated embedding calculation converges rapidly with respect to the number of border atoms.

Although the algorithm described in this section works well for relatively compact AO basis sets, such as the cc-pVDZ basis set used in all calculations up this point, it exhibits convergence failure for calculations that employ more diffuse basis sets, such as the aug-cc-pVDZ basis set. This is due to the well-known problem of charge leakage, in which the neglect of repulsive interactions in an embedding calculation allows for the improper transfer of electron density from the embedded subsystem to the surrounding environment.[51–53] As we show in Sec. III C, this problem can be remedied in the context of truncated projection-based embedding.

TABLE I. List of water molecules, the atoms of which comprise the set of border atoms for each value of R_{O-O} in Fig. 3(c). At $R_{O-O} = 3.0$ Å, the set of border atoms is the same as that shown in Fig. 1(b).

R_{O-O} (Å)	Molecules
2.0	
3.0	2, 3
4.2	2, 3, 4
5.0	2, 3, 4, 5
6.0	2, 3, 4, 5, 6

C. Switching between orbital projection and approximation of the non-additive kinetic potential (NAKP)

To address the problem of charge leakage in truncated embedding calculations employing diffuse basis sets, we include a simple modification to the truncated embedding algorithm from Sec. III B. Because that algorithm does not fully enforce mutual orthogonality between the subsystem A MOs and the MOs in $\{\phi_i\}_{B''}$, the NAKE between the corresponding electronic densities is non-zero. Accounting for this NAKE contribution requires modification of the embedded core Hamiltonian in Eq. (13), such that

$$\mathbf{h}^{A\text{ in }B} \approx \mathbf{h} + \mathbf{g}[\gamma^A + \gamma^B] - \mathbf{g}[\gamma^A] + \mu'\mathbf{P}^{B'} + \mathbf{v}^A_{NAKP}[\gamma^A, \gamma^{B''}], \quad (16)$$

where $\gamma^{B''}$ is the density matrix corresponding to the subsystem B MOs in $\{\phi_i\}_{B''}$, and the NAKP is

$$\mathbf{v}^A_{NAKP}[\gamma^A, \gamma^{B''}] = \mathbf{v}_{T_s}[\gamma^A + \gamma^{B''}] - \mathbf{v}_{T_s}[\gamma^A]. \quad (17)$$

The corresponding SCF-in-SCF energy from Eq. (4) is then

$$E_{SCF}[\gamma^A_{emb}; \gamma^A, \gamma^B] \approx E_{SCF}\left[\gamma^A_{emb}\right] + E_{SCF}[\gamma^B] + E^{nad}_{SCF}[\gamma^A, \gamma^B] + T^{nad}_s[\gamma^A, \gamma^{B''}] + \mathrm{tr}\left[\left(\gamma^A_{emb} - \gamma^A\right)\left(\mathbf{h}^{A\text{ in }B} - \mathbf{h}\right)\right], \quad (18)$$

where

$$T^{nad}_s[\gamma^A, \gamma^{B''}] = T_s[\gamma^A + \gamma^{B''}] - T_s[\gamma^A] - T_s[\gamma^{B''}], \quad (19)$$

and the corresponding WFT-in-SCF energy from Eq. (9) is

$$E_{WFT}[\Psi^A; \gamma^A, \gamma^B] \approx \langle\Psi^A|\hat{H}^{A\text{ in }B}|\Psi^A\rangle + E_{SCF}[\gamma^B] + E^{nad}_{SCF}[\gamma^A, \gamma^B] + T^{nad}_s[\gamma^A, \gamma^{B''}] - \mathrm{tr}[\gamma^A(\mathbf{h}^{A\text{ in }B} - \mathbf{h})]. \quad (20)$$

By construction, the overlap between the MOs in subsystem A and $\{\phi_i\}_{B''}$ is small; it can thus be expected that currently available approximations to the kinetic energy functional will provide an adequate description of the NAKE.

If all atoms are included in either the set of active or border atoms and if τ is sufficiently small, this approach corresponds to supermolecular projection-based embedding and involves no approximate kinetic energy (KE) functionals. In the other extreme, if no atoms are included in the set of border atoms, then no MOs are projected and the approach corresponds to the familiar case of monomolecular DFT embedding with the use of an approximate KE functional. The protocol in Eqs. (16)–(19) thus allows for the systematic switching between monomolecular DFT embedding and projection-based supermolecular embedding via modulation of τ and the set of border atoms.

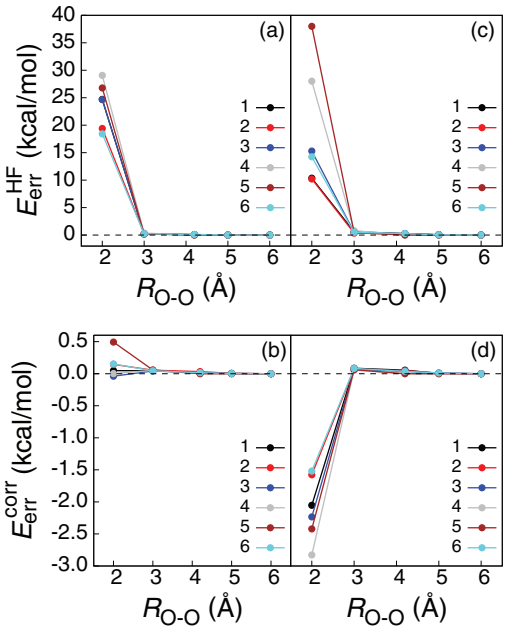

FIG. 4. (a) The effect of varying R_{O-O} on the HF-in-HF/cc-pVDZ embedding energy of the BK-1 conformation of the water hexamer. Each curve corresponds to assigning the constituent atoms of the indicated molecule as the set of active atoms. For a cutoff of 2.0 Å, the calculation is equivalent to a monomolecular calculation using TF embedding and no projection operator. At 6.0 Å, all of the calculations are performed in the supermolecular basis, and the projection operator is used exclusively with no approximate functionals. (b) The corresponding CCSD(T)-in-HF/cc-pVDZ results. (c) The corresponding HF-in-HF/aug-cc-pVDZ results. (d) The corresponding CCSD(T)-in-HF/aug-cc-pVDZ results.

To demonstrate this switching, Fig. 4 presents a series of truncated embedding calculations on the BK-1 water hexamer using the cc-pVDZ basis set. In each calculation, the active atoms correspond to one of the water molecules, $\{\mu', \tau\} = \{10^6, 0.5\}$, and $\mathbf{v}^{A}_{NAKP}[\gamma^{A}, \gamma^{B''}]$ is obtained using the Thomas-Fermi (TF) functional;[54,55] the border atoms are determined using a range of R_{O-O}. Figs. 4(a) and 4(b) show the effect of truncation in the HF-in-HF embedding calculations and in the CCSD(T)-in-HF embedding calculations, respectively. In both cases, the results are seen to quickly converge with respect to the number of border atoms.

In Figs. 4(c) and 4(d), these calculations are repeated using the larger aug-cc-pVDZ basis set. Again, the results converge rapidly with respect to the number of border atoms. However, the results in Figs. 4(c) and 4(d) contrast with those discussed in Sec. III B, for which truncated embedding with the larger basis set failed due to charge leakage. We thus find that inclusion of the NAKP between the subsystem A MOs and the MOs in $\{\phi_i\}_{B''}$ helps to mitigate the issue of charge leakage when basis set truncation is employed. This finding is consistent with earlier observations that monomolecular DFT embedding is a useful strategy for mitigating charge leakage.[56–58]

Finally, we note that Eqs. (16) and (18) can be regarded as a pairwise approximation,[31,32] such that

$$T_s^{nad}[\gamma^{A}, \gamma^{B'} + \gamma^{B''}] \approx T_s^{nad}[\gamma^{A}, \gamma^{B'}] + T_s^{nad}[\gamma^{A}, \gamma^{B''}]. \quad (21)$$

In the limit of $\mu' \to \infty$, the embedded subsystem MOs and the MOs in $\{\phi_i\}_{B'}$ are constrained to be mutually orthogonal for all γ^{A}; subject to this constraint, $T_s^{nad}[\gamma^{A}, \gamma^{B'}] = 0$ for all γ^{A}, and

$$\mathbf{v}^{A}_{NAKP}[\gamma^{A}, \gamma^{B'}] = \frac{\delta T_s^{nad}[\gamma^{A}, \gamma^{B'}]}{\delta \gamma^{A}} = 0. \quad (22)$$

Therefore, the only nonzero contribution to the NAKP is $\mathbf{v}^{A}_{NAKP}[\gamma^{A}, \gamma^{B''}]$ (Eq. (16)), and the only contribution to the NAKE is $T_s^{nad}[\gamma^{A}, \gamma^{B''}]$ (Eq. (18)).

IV. APPLICATIONS

A. WFT-in-HF truncated embedding for polypeptides

For a more demanding illustration of the truncated embedding approach presented in Sec. III C, we consider the Gly-Gly-Gly-Gly tetrapeptide. The optimized geometry of the tetrapeptide is determined at the HF/cc-pVDZ level, with all backbone dihedral angles constrained to 180°. For each of the truncated embedding calculations in this section, the set of active atoms consists of the atoms of one of the four glycine residues. The set of border atoms for each truncated embedding calculation is specified by a cutoff, n_t. If a backbone atom is within n_t bonds of an active atom, then it is included in the set of border atoms; if a non-backbone moiety (i.e., H, O, or OH) is bonded to a border atom, then its associated atoms are likewise included in the set of border atoms. Several sets of border atoms, each corresponding to a different value of n_t, are illustrated in Fig. 5 for the case in which the atoms of the Gly2 residue comprise the set of active atoms.

Fig. 6 illustrates that WFT-in-HF truncated embedding calculations on this system exhibit significant τ-dependence, since localization of the HF MOs yields orbitals with significant population on two or more backbone atoms. These calculations are performed using MP2-in-HF/aug-cc-pVDZ embedding, with the set of active atoms comprised of those in the Gly2 residue, with the set of border atoms associated with $n_t = 3$, and with $\mathbf{v}^{A}_{NAKP}[\gamma^{A}, \gamma^{B''}]$ obtained using the TF functional. Fig. 6(a) shows the number of projected MOs for several values of τ, and the μ'-dependence for each value of τ is shown in Fig. 6(b). For $\tau \geq 0.02$, there is very little μ'-dependence, whereas smaller values of τ lead to greater dependence on the level-shift parameter.

In general, it is preferable to set τ as small as possible without introducing significant μ'-dependence, since this results in fewer orbitals being treated at the level of the

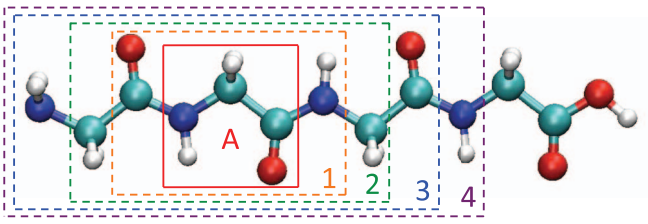

FIG. 5. The Gly-Gly-Gly-Gly tetrapeptide, with the set of active atoms comprised of the Gly2 residue (solid red box). Each of the dashed boxes indicates the union of the sets of active and border atoms for the corresponding value of n_t; any atoms outside of the boxes are included in the set of distant atoms.

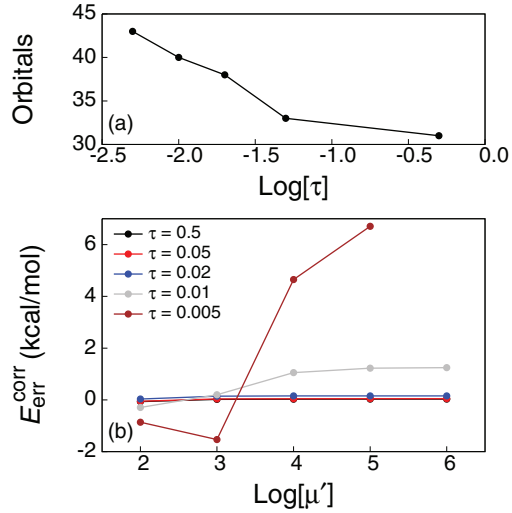

FIG. 6. (a) τ-dependence of the number of projected orbitals within MP2-in-HF/aug-cc-pVDZ truncated embedding calculations on the Gly-Gly-Gly-Gly tetrapeptide with $n_t = 3$. The choice of active and border atoms is indicated in Fig. 5. (b) μ'-dependence of the truncation error of this calculation for several values of τ.

approximate KE functional. For all systems considered in this paper, we find that $\tau = 0.05$ results in small μ'-dependencies; all remaining calculations reported in this paper thus employ $\{\mu', \tau\} = \{10^6, 0.05\}$ and utilize the TF functional to approximate $\mathbf{v}_{\mathrm{NAKP}}^{\mathrm{A}}[\gamma^{\mathrm{A}}, \gamma^{\mathrm{B''}}]$.

Fig. 7 presents additional MP2-in-HF embedding calculations using different sets of active atoms and using a range of values for the border atom cutoff, n_t. The set of active atoms associated with each curve corresponds to a different residue in the tetrapeptide. The results in Figs. 7(a) and 7(b) are obtained using the cc-pVDZ and aug-cc-pVDZ basis sets, respectively. Both sets of results converge rapidly with respect to the number of border atoms, although it is clear that a minimum of $n_t = 2$ is needed for the calculations with the aug-cc-pVDZ basis set; more diffuse basis functions in the augmented basis set lead to greater overlap between the subsystem A MOs and the MOs in $\{\phi_i\}_{\mathrm{B''}}$, thus increasing the contribution from the approximate NAKP functional and yielding a stronger dependence on the border atom cutoff.

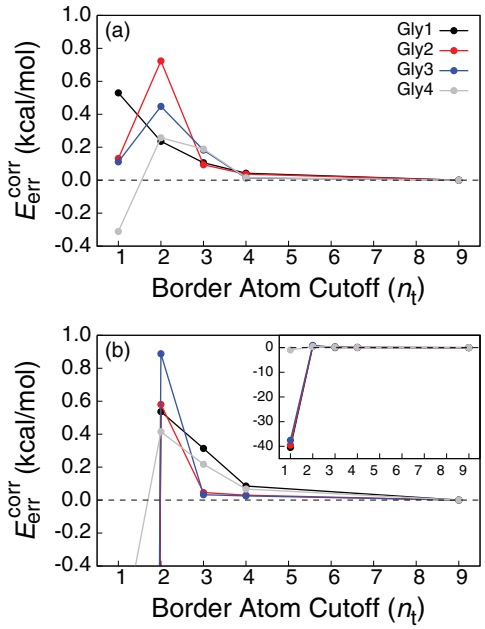

FIG. 7. (a) Convergence of the truncation error of embedding calculations on the Gly-Gly-Gly-Gly tetrapeptide using the cc-pVDZ basis set and several values of n_t. In each curve, the set of active atoms corresponds to the indicated residue. For $n_t = 9$, there are no distant atoms in any of the calculations. (b) The corresponding calculation using the aug-cc-pVDZ basis set. The inset shows the same results on a larger scale.

B. Embedded MBE

A promising application domain for projection-based WFT-in-HF embedding is the accurate MBE calculation of WFT energies.[30] This approach has the advantage of avoiding many of the challenges of more traditional MBE methods,[9,59–73] including sensitivity to the parameterization of point charges[74] or the need for "cap-atom" approximations.[75–80] As described previously, we perform the EMBE expansion in the correlation energy;[30] inclusion of the 1-body and 2-body terms yields the EMBE2 expression

$$E^{\mathrm{EMBE2}} = \sum_i E_i^{\mathrm{corr}} + \sum_{i>j} \left(E_{ij}^{\mathrm{corr}} - E_i^{\mathrm{corr}} - E_j^{\mathrm{corr}}\right), \quad (23)$$

where E_i^{corr} is the WFT-in-HF correlation energy of monomer i and E_{ij}^{corr} is the WFT-in-HF correlation energy of the dimer ij.

1. Water hexamers

EMBE2 calculations at the CCSD(T)-in-HF level are performed on a test set of 11 conformations of the water hexamer, using the 6-31G,[81,82] cc-pVDZ, and aug-cc-pVDZ basis sets. The calculations are performed with $\{\mu', \tau\} = \{10^6, 0.05\}$, and $\mathbf{v}_{\mathrm{NAKP}}^{\mathrm{A}}[\gamma^{\mathrm{A}}, \gamma^{\mathrm{B''}}]$ is obtained using the TF functional. Three of the hexamer geometries are taken from Ref. 83 and correspond to the (1) book, (2) cage, and (3) prism conformations; the other eight are taken from Ref. 45 and correspond to the (4) cyclic boat-1, (5) cyclic boat-2, (6) cyclic chair, (7) cage, (8) book-1, (9) book-2, (10) bag, and (11) an additional prism conformation. This test set includes a mixture of planar (Conf. 4, 5, and 6), quasi-planar (Conf. 1, 8, and 9), and three-dimensional (Conf. 2, 3, 7, and 11) conformations. Each monomer in the EMBE2 calculations corresponds to a set of active atoms comprised of one of the water molecules.

The relative energy of the water hexamer conformations are provided in Fig. 8(a), obtained using supermolecular EMBE2 calculations; full CCSD(T) calculations are also reported for comparison. Energies are reported with respect to that of Conf. (11), obtained using the corresponding level of theory and basis set. Fig. 8(b) presents the MBE error for each calculation, obtained using

$$E_{\mathrm{err}}^{\mathrm{MBE}} \equiv E^{\mathrm{EMBE2}} - E^{\mathrm{corr}}. \quad (24)$$

The mean unsigned MBE error, $\langle |E_{\mathrm{err}}^{\mathrm{MBE}}| \rangle$, of the EMBE2 calculations performed on this test set is 0.10 kcal/mol for the 6-31G basis set, 0.12 kcal/mol for the cc-pVDZ basis set, and

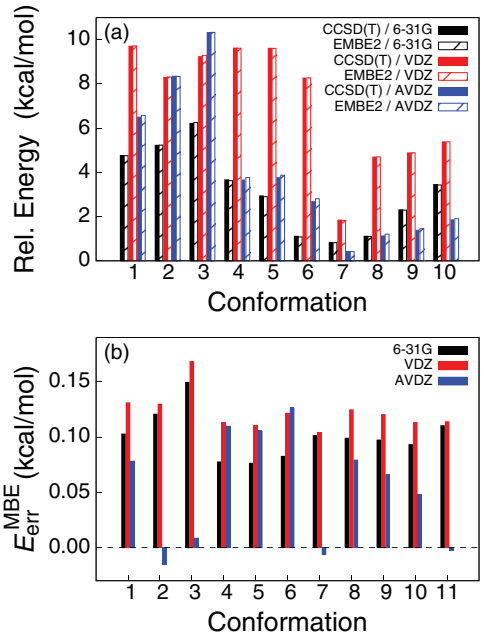

FIG. 8. (a) Energies of water hexamer conformations obtained using both CCSD(T) over the full system and CCSD(T)-in-HF supermolecular EMBE2 calculations. Three different basis sets are employed, with the cc-pVDZ and aug-cc-pVDZ basis sets abbreviated as VDZ and AVDZ, respectively. Conformation energies are reported with respect to the corresponding calculation for Conf. 11. (b) Error in the energy of the EMBE2 calculations.

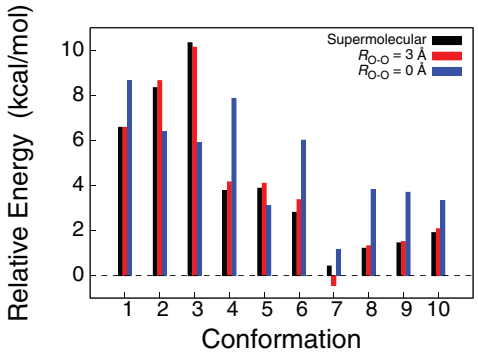

FIG. 9. Energies of water hexamer conformations obtained using both CCSD(T)-in-HF supermolecular EMBE2 calculations and CCSD(T)-in-HF truncated EMBE2 calculations. The embedding calculations employ truncated embedding with a border atom cutoff of either $R_{O-O} = 0$ Å or $R_{O-O} = 3$ Å. Conformation energies are reported with respect to the corresponding calculation for Conf. 11.

0.06 kcal/mol for aug-cc-pVDZ. The EMBE2 calculations are thus seen to produce smaller values of $\langle|E_{err}^{MBE}|\rangle$ than similar calculations using point-charge embedding;[68] equally important, however, is the fact that the embedding approach provided here rigorously avoids the problem of charge leakage, avoids the use of arbitrary parameters, and allows for full basis set convergence.

Fig. 9 presents the relative energies for the corresponding truncated embedding calculations in the aug-cc-pVDZ basis set. The border atoms are determined in the manner described in Sec. III B, using both $R_{O-O} = 0$ Å (i.e., monomolecular DFT embedding using the TF KE functional) and $R_{O-O} = 3$ Å. The truncated embedding calculations with $R_{O-O} = 3$ Å are in far better agreement with the reference supermolecular calculations, thus illustrating the potential of using truncated projection-based embedding to significantly improve upon the accuracy of DFT embedding with approximate KE functionals.

Table II presents a summary of the EMBE2 results for all the three basis sets (6-31G, cc-pVDZ, and aug-cc-pVDZ). The truncated embedding results using a non-empty set of border atoms (i.e., $R_{O-O} > 0$ Å) consistently produce smaller mean unsigned MBE errors than those obtained in the monomolecular basis (i.e., $R_{O-O} = 0$ Å). The standard deviation of the errors for each set of calculations, $\sigma[E_{err}^{MBE}]$, is also provided; this quantity reports on errors in the relative conformation energies, which may be of greater practical relevance than $\langle|E_{err}^{MBE}|\rangle$. For the embedding calculations employing a cutoff of $R_{O-O} = 0$ Å, $\sigma[E_{err}^{MBE}]$ exceeds 1 kcal/mol when the correlation-consistent basis sets are employed. For the embedding calculations employing a cutoff of $R_{O-O} = 3$ Å, $\sigma[E_{err}^{MBE}]$ is approximately 0.4 kcal/mol for each basis set, which is significantly smaller than the errors associated with the finite size of the basis sets (Fig. 8). Furthermore, the greater consistency of $\sigma[E_{err}^{MBE}]$ across the three basis sets for calculations that employ $R_{O-O} = 3$ Å rather than $R_{O-O} = 0$ Å indicates that truncated projection-based embedding provides more consistent errors in the relative energies than DFT embedding with approximate KE functionals. Finally, we note that in the limit of large R_{O-O} (i.e., supermolecular projection-based embedding) the precision of the results is further improved, in agreement with the expectation of controllable accuracy with respect to the choice of embedding parameters.

2. Polypeptides

EMBE2 calculations at the MP2-in-HF level are performed on several conformations of the Gly-Gly-Gly tripeptide using the cc-pVDZ and aug-cc-pVDZ basis sets. The calculations are performed with $\{\mu', \tau\} = \{10^6, 0.05\}$, and $\mathbf{v}_{NAKP}^{A}[\gamma^{A}, \gamma^{B''}]$ is obtained using the TF functional. The geometries are obtained via optimization at the HF/cc-pVDZ

TABLE II. Summary of the EMBE2 results for the water hexamer test set. Results are listed using truncated embedding with a cutoff of $R_{O-O} = 0$ Å, truncated embedding with a cutoff of $R_{O-O} = 3$ Å, and supermolecular embedding. All values are in kcal/mol.

	$\langle\|E_{err}^{MBE}\|\rangle$		
Truncation level	6-31G	cc-pVDZ	aug-cc-pVDZ
$R_{O-O} = 0$ Å	0.54	3.40	6.25
$R_{O-O} = 3$ Å	0.39	1.03	0.95
Supermolecular	0.10	0.12	0.06
	$\sigma[E_{err}^{MBE}]$		
Truncation level	6-31G	cc-pVDZ	aug-cc-pVDZ
$R_{O-O} = 0$ Å	0.35	6.38	2.53
$R_{O-O} = 3$ Å	0.42	0.40	0.40
Supermolecular	0.02	0.02	0.05

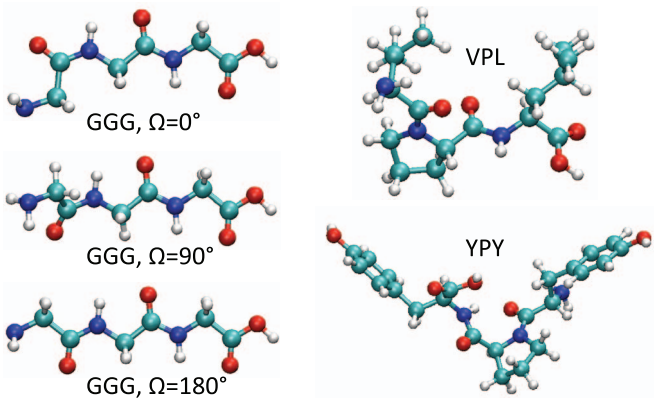

FIG. 10. Three of the Gly-Gly-Gly (GGG) tripeptide conformations are presented on the left for several different dihedral angles. The geometries of the Val-Pro-Leu (YPL) and Tyr-Pro-Tyr (YPY) tripeptides are presented on the right.

level, with the Gly1-Gly2 bond dihedral (Ω) constrained to several values, and with all other backbone dihedral angles constrained to 180°. Several of these geometries are shown at left in Fig. 10. Each monomer in the EMBE2 calculations corresponds to a set of active atoms comprised of one of the tripeptide residues. The sets of border atoms employed in the EMBE2 calculations are defined in terms of the n_t cutoff, as described in Sec. IV A.

Figs. 11(a) and 11(b) present the correlation energies from EMBE2 calculations on the Gly-Gly-Gly tripeptide conformations using the cc-pVDZ and aug-cc-pVDZ basis sets, respectively; each correlation energy is reported relative to that of the corresponding calculation on the conformation

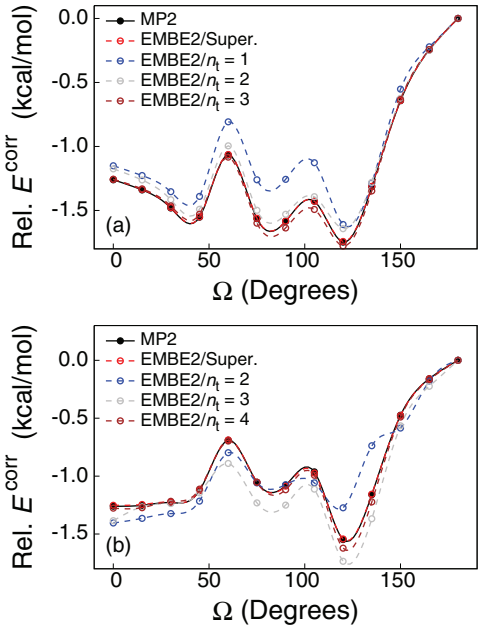

FIG. 11. (a) Gly-Gly-Gly tripeptide conformation energies obtained using MP2-in-HF EMBE2 calculations and employing the cc-pVDZ basis. Conformation energies are reported with respect to the corresponding calculation for the Ω = 180° conformation. The results using $n_t = 4$ are not shown for this basis set, as they are nearly indistinguishable from the supermolecular results. (b) The corresponding results employing the aug-cc-pVDZ basis. The results using $n_t = 1$ are not shown for this basis set, as they are highly inaccurate.

with Ω = 180°. It is seen that the truncated embedding calculations reproduce the trends in the relative energies of the reference MP2 calculations, and that the accuracy improves with the number of border atoms. Table III lists the corresponding values of $\langle|E_{\rm err}^{\rm MBE}|\rangle$ and $\sigma[E_{\rm err}^{\rm MBE}]$. In agreement with the results in Fig. 7, sets of border atoms associated with $n_t \geq 2$ are needed to achieve suitable accuracy with the aug-cc-pVDZ basis set. Both $\langle|E_{\rm err}^{\rm MBE}|\rangle$ and $\sigma[E_{\rm err}^{\rm MBE}]$ are generally found to improve with increasing numbers of border atoms. These results demonstrate that truncated EMBE2 calculations yield accurate results for systems in which embedding is performed across covalently bound monomers. Furthermore, since Ω is associated with rotation of a bond that connects different monomers, these results indicate that the EMBE2 calculations are relatively robust with respect to changes in the electronic environment in the inter-subsystem covalent bonds.

To illustrate the corresponding calculations for tripeptides with different side-chains, additional EMBE2 calculations are performed on the Val-Pro-Leu and Tyr-Pro-Tyr tripeptides. These tripeptides include both hydrophobic and hydrophilic side-chains, including residues with aromatic rings; in particular, we note that the proline side-chains present an interesting challenge to the accuracy of the truncated embedding calculations, since they exhibit covalent bonds to multiple backbone atoms. Geometries for these tripeptides are optimized at the HF/cc-pVDZ level of theory, with the initial position of the heavy atoms obtained from reported crystal structures (Fig. 10, right).[84,85] For the truncated embedding calculations, the atoms of side-chain moieties are only included in the set of border atoms if all backbone atoms to which the side-chain moieties are bonded are border atoms.

Table IV presents the results of EMBE2 calculations for the Val-Pro-Leu and Tyr-Pro-Tyr tripeptides, as well as for the Gly-Gly-Gly-Gly tetrapeptide from Sec. IV A. Due to the computational cost of the reference calculations, results employing the aug-cc-pVDZ basis set are not included for these more complex tripeptides. As with the Gly-Gly-Gly tripeptide calculations (Fig. 11 and Table III), the results yield small (sub kcal/mol) errors that systematically decrease with the number of border atoms.

TABLE III. Summary of the EMBE2 results for the Gly-Gly-Gly tripeptide. All calculations use either the cc-pVDZ (VDZ) basis set or the aug-cc-pVDZ (AVDZ) basis set. Results are provided for several values of n_t, as well as for the supermolecular basis set (Super.). Both the mean unsigned MBE error over all values of Ω and the standard deviation of the MBE error are provided. All values are reported in kcal/mol.

	$\langle\|E_{\rm err}^{\rm MBE}\|\rangle$				
Basis	$n_t = 1$	$n_t = 2$	$n_t = 3$	$n_t = 4$	Super.
VDZ	0.106	0.345	0.103	0.007	0.050
AVDZ	18.549	0.678	0.216	0.054	0.056
	$\sigma[E_{\rm err}^{\rm MBE}]$				
Basis	$n_t = 1$	$n_t = 2$	$n_t = 3$	$n_t = 4$	Super.
VDZ	0.112	0.033	0.028	0.008	0.002
AVDZ	19.148	0.167	0.080	0.026	0.005

TABLE IV. The MBE error (Eq. (24)) for the Val-Pro-Lue tripeptide, the Tyr-Pro-Tyr tripeptide, and the Gly-Gly-Gly-Gly tetrapeptide EMBE2 calculations. All calculations use either the cc-pVDZ (VDZ) basis set or the aug-cc-pVDZ (AVDZ) basis set. Results are provided for several values of n_t, as well as for the supermolecular basis set (Super.). All values are reported in kcal/mol.

Peptide/basis	$n_t = 1$	$n_t = 2$	$n_t = 3$	$n_t = 4$	Super.
VPL/VDZ	− 0.029	− 1.041	− 0.413	− 0.205	0.037
YPY/VDZ	0.604	− 0.767	− 0.092	− 0.277	0.076
GGGG/VDZ	− 0.175	− 0.821	− 0.751	− 0.066	0.095
GGGG/AVDZ	26.439	− 1.513	− 1.118	− 0.234	0.108

V. CONCLUSIONS

In this paper, we have presented an extension of our projection-based embedding method to allow for truncation of the AO basis set for subsystem calculations. The truncation approach involves combining highly accurate projection-based embedding for nearby interactions with an approximate treatment of the NAKP between distant MOs. Application of this approach to both molecular clusters and polypeptides illustrates that the errors introduced by truncation of the AO basis set are both small and systematically controllable with respect to the extent of truncation. EMBE calculations on these systems yield accurate total and relative conformational energies, even when the monomers in the expansion are connected by covalent bonds. Furthermore, we have demonstrated that this approach offers a means of switching between accurate projection-based embedding and DFT embedding using approximate KE functionals, such that it both benefits from previous research on the development of approximate KE functionals and allows for systematic improvement upon those functionals in practical applications. These results establish that the projection-based embedding method enables efficient WFT-in-SCF embedding calculations on large molecular systems.

ACKNOWLEDGMENTS

This work is supported by the U. S. Army Research Laboratory and the U. S. Army Research Office (USARO) under Grant No. W911NF-10-1-0202, by the Air Force Office of Scientific Research (USAFOSR) under Grant No. FA9550-11-1-0288, and by the Office of Naval Research (ONR) under Grant No. N00014-10-1-0884. T.A.B. acknowledges support from a National Defense Science and Engineering Graduate Fellowship, and T.F.M. acknowledges support from a Camille and Henry Dreyfus Foundation New Faculty Award and an Alfred P. Sloan Foundation Research Fellowship.

[1] A. Warshel and M. Levitt, J. Mol. Biol. **103**, 227 (1976).
[2] P. Sherwood, A. H. de Vries, S. J. Collins, S. P. Greatbanks, N. A. Burton, M. A. Vincent, and I. H. Hillier, Faraday Discuss. **106**, 79 (1997).
[3] J. L. Gao, P. Amara, C. Alhambra, and M. J. Field, J. Phys. Chem. A **102**, 4714 (1998).
[4] H. Lin and D. G. Truhlar, Theor. Chem. Acc. **117**, 185 (2007).
[5] H. M. Senn and W. Thiel, Angew. Chem., Int. Ed. **48**, 1198 (2009).
[6] L. Hu, P. Söderhjelm, and U. Ryde, J. Chem. Theory Comput. **7**, 761 (2011).
[7] S. Dapprich, I. Komáromi, K. S. Byun, K. Morokuma, and M. J. Frisch, THEOCHEM **461–462**, 1 (1999).
[8] F. Maseras and K. Morokuma, J. Comput. Chem. **16**, 1170 (1995).
[9] K. Kitaura, E. Ikeo, T. Asada, T. Nakano, and M. Uebayasi, Chem. Phys. Lett. **313**, 701 (1999).
[10] D. G. Fedorov and K. Kitaura, J. Chem. Phys. **120**, 6832 (2004).
[11] D. G. Fedorov and K. Kitaura, J. Phys. Chem. A **111**, 6904 (2007).
[12] P. Arora, W. Li, P. Piecuch, J. W. Evans, M. Albao, and M. S. Gordon, J. Phys. Chem. C **114**, 12649 (2010).
[13] S. R. Pruitt, M. A. Addicoat, M. A. Collins, and M. S. Gordon, Phys. Chem. Chem. Phys. **14**, 7752 (2012).
[14] K. R. Brorsen, N. Minezawa, F. Xu, T. L. Windus, and M. S. Gordon, J. Chem. Theory Comput. **8**, 5008 (2012).
[15] A. Gaenko, T. L. Windus, M. Sosonkina, and M. S. Gordon, J. Chem. Theory Comput. **9**, 222 (2013).
[16] N. Govind, Y. A. Wang, A. J. R. da Silva, and E. A. Carter, Chem. Phys. Lett. **295**, 129 (1998).
[17] N. Govind, Y. A. Wang, and E. A. Carter, J. Chem. Phys. **110**, 7677 (1999).
[18] T. Klüner, N. Govind, Y. A. Wang, and E. A. Carter, J. Chem. Phys. **116**, 42 (2002).
[19] P. Huang and E. A. Carter, J. Chem. Phys. **125**, 084102 (2006).
[20] S. Sharifzadeh, P. Huang, and E. Carter, J. Phys. Chem. C **112**, 4649 (2008).
[21] A. S. P. Gomes, C. R. Jacob, and L. Visscher, Phys. Chem. Chem. Phys. **10**, 5353 (2008).
[22] T. A. Wesolowski, Phys. Rev. A **77**, 012504 (2008).
[23] Y. G. Khait and M. R. Hoffmann, J. Chem. Phys. **133**, 044107 (2010).
[24] C. Huang, M. Pavone, and E. A. Carter, J. Chem. Phys. **134**, 154110 (2011).
[25] C. Huang and E. A. Carter, J. Chem. Phys. **135**, 194104 (2011).
[26] S. Hofener, A. S. P. Gomes, and L. Visscher, J. Chem. Phys. **136**, 044104 (2012).
[27] O. Roncero, A. Zanchet, P. Villarreal, and A. Aguado, J. Chem. Phys. **131**, 234110 (2009).
[28] A. S. P. Gomes and C. R. Jacob, Annu. Rep. Prog. Chem., Sect. C: Phys. Chem. **108**, 222 (2012).
[29] J. D. Goodpaster, T. A. Barnes, F. R. Manby, and T. F. Miller III, J. Chem. Phys. **137**, 224113 (2012).
[30] F. R. Manby, M. Stella, J. D. Goodpaster, and T. F. Miller III, J. Chem. Theory Comput. **8**, 2564 (2012).
[31] J. D. Goodpaster, N. Ananth, F. R. Manby, and T. F. Miller III, J. Chem. Phys. **133**, 084103 (2010).
[32] J. D. Goodpaster, T. A. Barnes, and T. F. Miller III, J. Chem. Phys. **134**, 164108 (2011).
[33] S. Fux, C. R. Jacob, J. Neugebauer, L. Visscher, and M. Reiher, J. Chem. Phys. **132**, 164101 (2010).
[34] J. Nafziger, Q. Wu, and A. Wasserman, J. Chem. Phys. **135**, 234101 (2011).
[35] P. D. Dedíková, P. Neogrády, and M. Urban, J. Phys. Chem. A **115**, 2350 (2011).
[36] P. G. Lykos and R. G. Parr, J. Chem. Phys. **24**, 1166 (1956).
[37] J. C. Phillips and L. Kleinman, Phys. Rev. **116**, 287 (1959).
[38] H. Stoll, B. Paulus, and P. Fulde, J. Chem. Phys. **123**, 144108 (2005).
[39] R. A. Mata, H.-J. Werner, and M. Schütz, J. Chem. Phys. **128**, 144106 (2008).
[40] T. M. Henderson, J. Chem. Phys. **125**, 014105 (2006).
[41] S. Huzinaga and A. A. Cantu, J. Chem. Phys. **55**, 5543 (1971).
[42] B. Swerts, L. F. Chibotaru, R. Lindh, L. Seijo, Z. Barandiaran, S. Clima, K. Pierloot, and M. F. A. Hendrickx, J. Chem. Theory Comput. **4**, 586 (2008).
[43] J. L. Pascual, N. Barros, Z. Barandiaran, and L. Seijo, J. Phys. Chem. A **113**, 12454 (2009).
[44] H.-J Werner, P. J. Knowles, R. Lindh, F. R. Manby, M. Shütz *et al.*, MOLPRO, version 2012.1, a package of *ab initio* programs, 2012, see www.molpro.net.
[45] B. Temelso, K. A. Archer, and G. C. Shields, J. Phys. Chem. A **115**, 12034 (2011).
[46] T. H. Dunning, Jr., J. Chem. Phys. **90**, 1007 (1989).
[47] R. A. Kendall, T. H. Dunning, Jr., and R. J. Harrison, J. Chem. Phys. **96**, 6796 (1992).
[48] See supplementary material at http://dx.doi.org/10.1063/1.4811112 for the employed molecular geometries.
[49] J. Pipek and P. G. Mezey, J. Chem. Phys. **90**, 4916 (1989).
[50] K. Raghavachari, G. W. Trucks, J. A. Pople, and M. Head-Gordon, Chem. Phys. Lett. **157**, 479 (1989).
[51] E. V. Stefanovich and T. N. Truong, J. Chem. Phys. **104**, 2946 (1996).
[52] A. Laio, J. VandeVondele, and U. Rothlisberger, J. Chem. Phys. **116**, 6941 (2002).
[53] K. Senthilkumar, J. I. Mujika, K. E. Ranaghan, F. R. Manby, A. J. Mulholland, and J. N. Harvey, J. R. Soc., Interface **5**, S207 (2008).

[54]L. H. Thomas, Proc. Cambridge Philos. Soc. **23**, 542 (1927).
[55]E. Fermi, Z. Phys. **48**, 73 (1928).
[56]C. R. Jacob, T. A. Wesolowski, and L. Visscher, J. Chem. Phys. **123**, 174104 (2005).
[57]M. Dulak and T. A. Wesolowski, J. Chem. Phys. **124**, 164101 (2006).
[58]C. R. Jacob, S. M. Beyhan, and L. Visscher, J. Chem. Phys. **126**, 234116 (2007).
[59]E. E. Dahlke and D. Truhlar, J. Phys. Chem. B **110**, 10595 (2006).
[60]S. Hirata, O. Sode, M. Keçeli, and T. Shimazaki, *Accurate Condensed Phase Quantum Chemistry* (Taylor and Francis, 2011).
[61]P. J. Bygrave, N. L. Allan, and F. R. Manby, J. Chem. Phys. **137**, 164102 (2012).
[62]C. R. Taylor, P. J. Bygrave, J. N. Hart, N. L. Allan, and F. R. Manby, Phys. Chem. Chem. Phys. **14**, 7739 (2012).
[63]E. E. Dahlke and D. G. Truhlar, J. Chem. Theory Comput. **3**, 46 (2007).
[64]E. E. Dahlke and D. G. Truhlar, J. Chem. Theory Comput. **3**, 1342 (2007).
[65]M. Isegawa, B. Wang, and D. G. Truhlar, J. Chem. Theory Comput. **9**, 1381 (2013).
[66]E. E. Dahlke and D. G. Truhlar, J. Chem. Theory Comput. **4**, 1 (2008).
[67]A. Sorkin, E. E. Dahlke, and D. G. Truhlar, J. Chem. Theory Comput. **4**, 683 (2008).
[68]E. E. Dahlke, H. R. Leverentz, and D. G. Truhlar, J. Chem. Theory Comput. **4**, 33 (2008).
[69]L. D. Jacobson and J. M. Herbert, J. Chem. Phys. **134**, 094118 (2011).
[70]S. Hirata, Phys. Chem. Chem. Phys. **11**, 8397 (2009).
[71]R. M. Richard and J. M. Herbert, J. Chem. Phys. **137**, 064113 (2012).
[72]G. J. O. Beran, J. Chem. Phys. **130**, 164115 (2009).
[73]S. Wen, K. Nanda, Y. Huang, and G. J. O. Beran, Phys. Chem. Chem. Phys. **14**, 7578 (2012).
[74]E. B. Kadossov, K. J. Gaskell, and M. A. Langell, J. Comput. Chem. **28**, 1240 (2007).
[75]U. C. Singh and P. A. Kollman, J. Comput. Chem. **7**, 718 (1986).
[76]M. J. Field, P. A. Bash, and M. Karplus, J. Comput. Chem. **11**, 700 (1990).
[77]V. Deev and M. A. Collins, J. Chem. Phys. **122**, 154102 (2005).
[78]M. A. Collins and V. A. Deev, J. Chem. Phys. **125**, 104104 (2006).
[79]D. W. Zhang, X. H. Chen, and J. Z. H. Zhang, J. Comput. Chem. **24**, 1846 (2003).
[80]D. W. Zhang and J. Z. H. Zhang, J. Chem. Phys. **119**, 3599 (2003).
[81]R. Ditchfield, W. J. Hehre, and J. A. Pople, J. Chem. Phys. **54**, 724 (1971).
[82]W. J. Hehre, R. Ditchfield, and J. A. Pople, J. Chem. Phys. **56**, 2257 (1972).
[83]M. D. Tissandier, S. J. Singer, and J. V. Coe, J. Phys. Chem. A **104**, 752 (2000).
[84]S. C. Graham and J. M. Guss, Arch. Biochem. Biophys. **469**, 200 (2008).
[85]Z. Zhang, M. Qian, Q. Huang, Y. Jia, Y. Tang, K. Wang, D. Cui, and M. Li, J. Protein Chem. **20**, 59 (2001).